Recent trends in co-authorship in economics: evidence from RePEc

Katharina Rath and Klaus Wohlrabe

Ifo Institute for Economic Research

17. August 2015

Recent trends in co-authorship in economics: evidence from RePEc

Katharina Rath
Ifo Institute – Leibniz-Institute
for Economic Research
at the University of Munich e.V.
Poschingerstr. 5
81679 Munich, Germany
Phone: +49(0)89/9224-1229
k.rath@campus.lmu.de

Klaus Wohlrabe*
Ifo Institute – Leibniz-Institute
for Economic Research
at the University of Munich e.V.
Poschingerstr. 5
81679 Munich, Germany
Phone: +49(0)89/9224-1229
wohlrabe@ifo.de

* Corresponding author.

Abstract

This articles investigates the recent trends in co-authorship in economics. Using data from more than 700.000 journal articles we show that the average number of authors per has increased over the last years. This process is likely to be continued in the future. In a regression analysis we present evidence how the authorship of papers is related to the number of citations, the JEL classification, the number of journal pages and the length of the title.

JEL Code: A12, A14
Keywords: co-authorship, economics, trends, RePEc

1 Introduction

It is a well-known phenomenon that the number of co-authors in economics has increased over the last years. This has been documented in several studies using rather small set of journals. Sutter and Kocher (2004) show that the share of co-authored papers in fifteen economics journals increased from under 30 % in 1977 to 54 % in 1997. A similar trend is reflected by McDowell and Melvin (1983) for eight economics journals, Laband and Tollison (2000) for three economics journals and Nowell and Grijalva (2011) for publications of faculty members of 129 US colleges. Additionally, the number of co-authors increased as most recently shown by Card and Della Vigna (2013) for five economics journals. Furthermore, Goyal, Van Der Leij, and Moraga-González (2006) show how the "social distance" between all authors in economics decreased between 1970 and 2000. The number of authors to link up all economists via co-authorship connections became lower over time. Kosnik (2015) shows, using data from nine economics journals, that articles of JEL categories with higher shares of co-authorship generally contain more keywords. In microeconomics and mathematical methods articles, co-authored papers have an increasing use of technical terms. Single-authored papers in these categories have a constant use of such terms over time.

Following closely Nowell and Grijalva (2011) (henceforth NG), we add to this literature in several ways. We employ the largest data set so far. We analyse 752,680 articles from 1615 journals over the time period from 1991 to 2013. Firstly, this allows to set the analysis on a broader basis and investigating recent trends. Furthermore we show how journal quality, length of the title and the article affects the number of co-authors. Additionally, we show how the co-authorship patterns of scientific rookies developed over time.

The remainder of the paper is as follows: after describing the data set from RePEc we provide some descriptive statistics of recent trends in authorship. Then we run a Poisson regression as in NG which uncovers the determinants of co-authorship. Finally, we conclude.

2 Data

We extracted our data from RePEc (Research Papers in Economics, www.repec.org). In economics, RePEc has become an essential source for the spread of knowledge and ranking of individual authors and academic institutions. RePEc is based on the "active participation principle", i.e. that authors, institutions and publishers have to register and to provide information to the network. This approach has the main advantage that a clear assignment of works and citations to authors and articles is possible. Indeed, the RePEc story has become a success, with more than 45,000 registered authors with listed works and 2,250 journals in economic sciences worldwide as of August 2015. Using a unique identifier, we downloaded all meta-information for more than 1,000,000 journal articles listed in RePEc. This includes the title, the journal, the beginning and ending page, number of authors and citations, as well as the JEL classifications. Furthermore, we were able to identify registered authors with RePEc and to assign their corresponding articles. As the current JEL classification started in 1991 we excluded all articles before 1991. Additionally, we restricted ourselves to data up to 2013, as for 2014 not all information for all journals were available. We also excluded articles with obviously misclassified bibliometric information. All data were downloaded on 01/19/2015. Finally, we have data for 752,680 journal articles.

The quality level of a journal is captured by the simple impact factor. The definition is similar to the "official" impact factor published by Thomson Reuters Journal Citation Reports. The main difference is the year and article coverage of citation counts. In RePEc, all citations are related to total number of registered articles in a journal. For further details on RePEc see Zimmermann (2013) and Seiler and Wohlrabe (2012).

Additionally, around 17% of the RePEc articles are research classified by Journal of Economic Literature (JEL) codes. We increase this share by including the JEL data from EconLit and the respective journal websites. We end up with 219,067 (30%) articles with JEL classifications.[1]

Furthermore, we are able to analyse individual time trends in co-authorship for authors

[1] JEL codes are not for all journals available. We excluded the Y and Z category from our analysis.

that are registered with RePEc. Each of these authors has a unique identifier, which allows us to match the respective research output. On this basis, we can derive the research activity period, which starts with year of the first publication. Our article set described above cannot be completely related to the registered authors. This leaves us for a sub-analysis with 354,928 articles.

3 A new look on trends of co-authorship in economics

In Figure 1, we show how the average number of authors per article changed over time. With an increase from 1.56 in 1991 to 2.23 in 2013 authors per paper, a clear upward trend is visible. As stated before we formed two sub-groups of articles: articles from registered authors and JEL-classified articles. The co-authorship development is also depicted in Figure 1. The upward trend is almost the same. The main difference is that registered authors tend to have on average more co-authors than the complement group. One reason might be that registered authors are better networkers. This confirms recent findings in the literature. Based on our large data set, we run a simple regression as in NG. We define a dummy variable with 1 if a paper is co-authored and 0 otherwise. As explanatory variables we employ a linear and quadratic time trend. In Table 1, we show the results of the linear probability model. In contrast to NG we find a significant positive quadratic trend. However, the economic significance is rather small. This implies that the trend of increasing share of co-authorship in economics continues over the next years. This result is confirmed using the number of authors as the dependent variable (right panel of Table 1). Based on these regression results, we have in 2020 a share of co-authored papers of 73% and an average of 2.55 authors per paper.

Table 1: Time trend regressions for co-authorship

Dependent Variable	Co-authored		Number of Authors	
	Estimated coefficient	p-value	Estimated coefficient	p-value
Constant	0.392	0.000	1.561	0.000
Year	0.010	0.000	0.012	0.000
$Year^2$	0.000	0.000	0.001	0.000
F (Significance)	7379.38 (0.00)		9875.88 (0.00)	

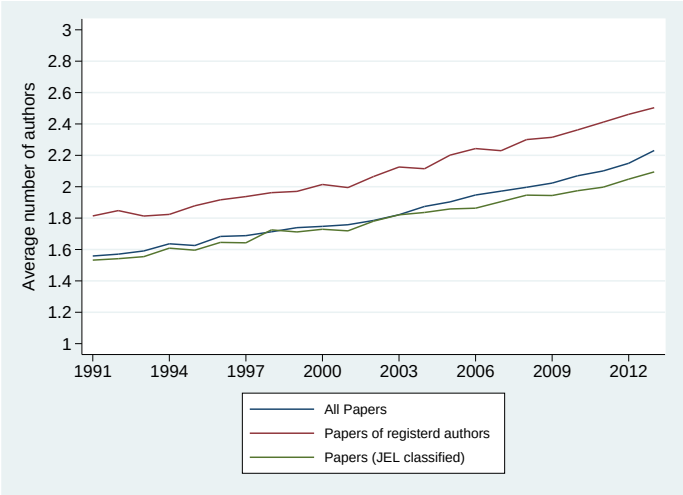

Figure 1: Time trend over average number of authors for all articles and two sub-groups

Figure 2 (left panel) plots the distribution of authorship. In total, 45% of papers were single-authored, 32% had two authors, and 23% had three or more authors. However, this result should be reflected in light of the time trend. This is done in right panel of Figure 2. The share of single-authored papers has fallen by 20 percentage points. This result is driven by a large increase of papers with three or more authors. Surprisingly, the share of papers with two authors has remained nearly constant.

Does the trend in authorship shown in Figure 1 hold across different journals quality classes? The answer is given in Figure 3 where the number of authors is again plotted over time. However, this time it is differentiated between journal quality using quartiles of RePEc impact factors. For all quality classes, the number of authors has increased over time. However, papers published in journals ranked in the lowest quality quartile have on average less authors. In 2013, the average number of authors in this category was 1.84, whereas for the other three it ranged between 2.29 and 2.51. This might be difficult to explain. One reason could be that the quality of an article increases with their number of authors. We leave this for future research.

Additionally, one might wonder if the increasing trend in authorship holds rather for established authors than for economists at the beginning of their publishing career. The beginning

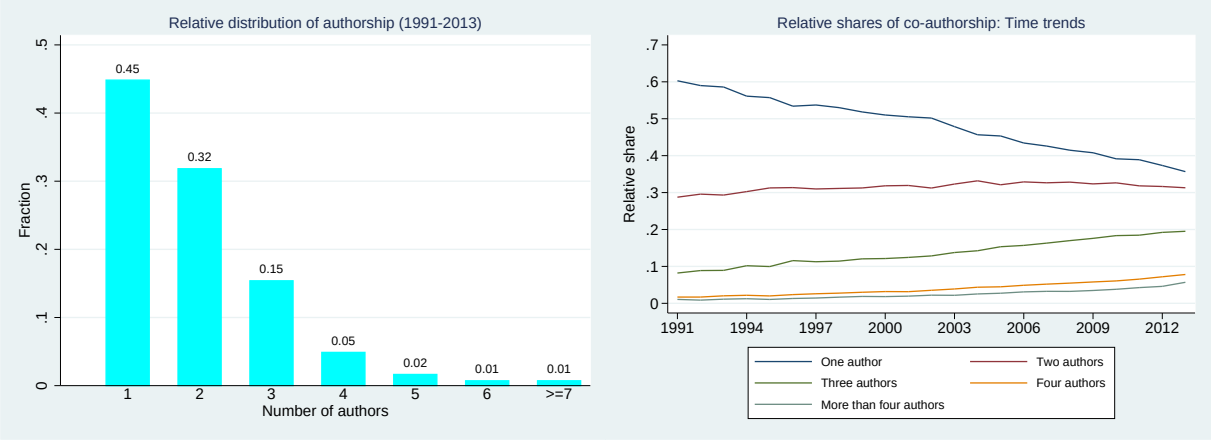

Figure 2: Relative distribution of authorship

of the career is defined by the year of the first journal article. Starting an academic career an economist should demonstrate that he or she is able to do research on their own, i.e. mainly without co-authors. The job market paper in economics is expected to be single-authored. Therefore, we presume that the average number of authors of papers by scientific rookies is smaller than the overall average. Figure 4 shows, that this is not the case. First-time authors work with slightly less co-authors. Nevertheless, the number of authors had increased as well for new authors.

In Figure 5, we plot the time trend of authorship for each JEL category separately. For all JEL codes we detect a clear upward trend.[2] The category with the smallest average number over the full time period is "History of Economic Thought, Methodology, and Heterodox Approaches" (B). In 2013 the category with the highest share was "Agricultural and Natural Resource Economics, Environmental and Ecological Economics" (Q).

4 Regression Analysis

In this section, we repeat the regression analysis of NG. We focus on the Poisson regression with the number of authors per paper as the dependent variable.[3] As independent variables

[2] We skipped the legend in the figure in order to save space.

[3] We also performed the logit regression of NG and obtained qualitatively the same results.

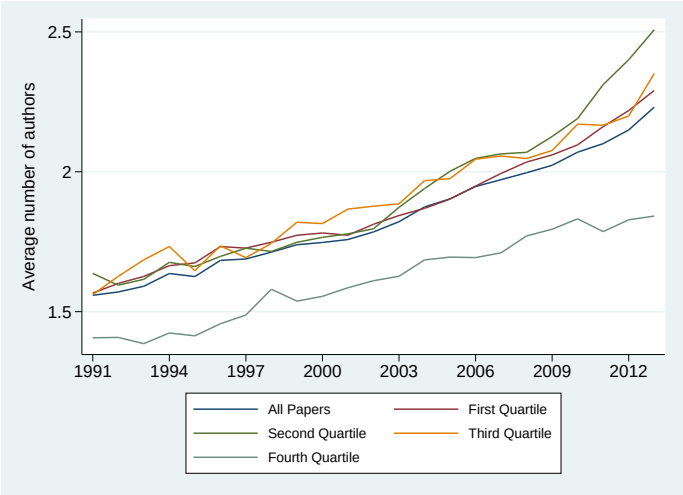

Figure 3: Average number of authors for quality quartiles

we included a time trend, a squared time trend, journal quality represented as the log of the RePEc impact factor, the number of JEL categories per paper and dummies for each JEL category. In addition to NG we also included the length the title counted as characters and the number of journal pages. In Table 2 we outline the descriptive statistics for each variable. In Table 3, we present the regression results for three different specifications. In Model 1, we included the linear and the quadratic time trend, the IF of a journal and the number of JEL codes per paper. All variables are significant and have the expected signs with the exception of the quadratic time trend. This is contrast to NG who found a negative quadratic time trend. In Model 2, we add the number of citations. This variable turns out to be highly significant. The higher the number of authors the higher is the visibility of an article. This might be reflected in more conference visits, publication in various working paper outlets and finally in high (self-) citation counts. In our Model 3, we also included the length of title of a paper, the number of pages and the dummies for all JEL categories. Again, almost all variables are significant. As it was to expected, the higher the number of authors the longer the journal article. What remains puzzling is that the length of the title is significantly positively associated with the number of co-authors. There is no obvious explanation for this observation. As in NG, we find that co-authorship patterns differs between sub-disciplines of

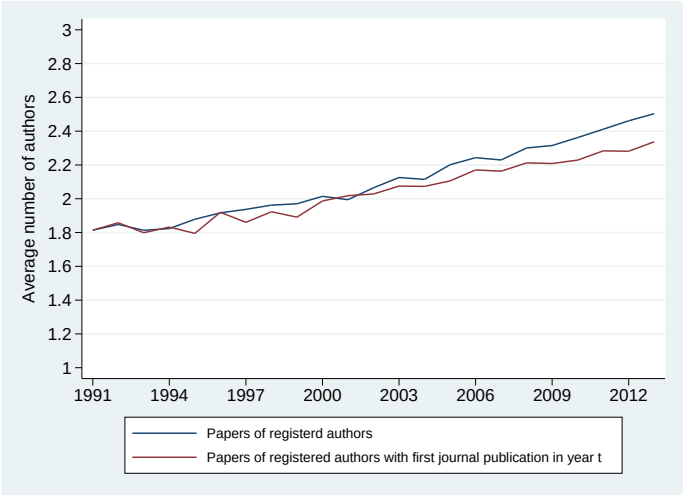

Figure 4: Comparison of the full sample and scientific rookies

economics. Our estimated coefficients for the JEL categories are significant in 16 out 18 cases. In contrast, NG find in their Model 3 in Table 5 only 11 out 17 cases to be significant.[4]

5 Conclusion

This article adds to the recent research on patterns of co-authorship in economics over time. Based on a large data set of more than 700,000 articles we verify the constantly increasing trend of authorship. We mainly confirm the hypotheses stated in Nowell and Grijalva (2011). The upward trend is visible across all sub-disciplines in economics and across different quality levels of journals. For the latter we find, that the higher the impact of a journal the higher the likelihood of more co-authors. The decrease of solo-authored papers over time is replaced by papers with more than two authors. The share of the latter one is relatively stable over time. We are first to show that the co-authorship patter of scientific rookies is following the overall trend found in this paper. In our regression results we find a clear positive relationship between co-authorship and the number of citations, the length of an article and the title. Especially the latter one is surprising. We leave this for future research.

[4]NG summarized the JEL categories B and N.

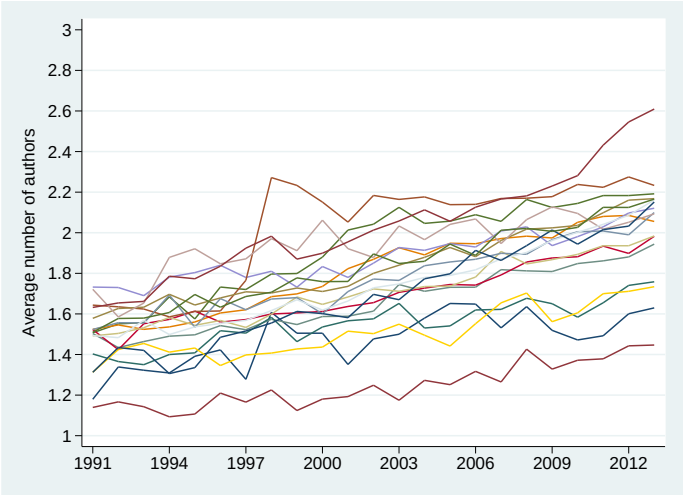

Figure 5: Time trend over aggregate number of authors for each JEL category

References

Goyal, S., M. J. Van Der Leij, and J. L. Moraga-González (2006): "Economics: An emerging small world," *Journal of Political Economy*, 114(2), 403–412.

Kosnik, L.-R. (2015): "What have economists been doing for the last 50 years? A text analysis of published academic research from 1960-2010," *Economics: The Open-Access, Open-Assessment E-Journal*, 9(2015-13), 1–38.

Laband, D. N., and R. D. Tollison (2000): "Intellectual collaboration," *Journal of Political Economy*, 108(3), 632–662.

McDowell, J. M., and M. Melvin (1983): "The determinants of co-authorship: An analysis of the economics literature," *The Review of Economics and Statistics*, 65(1), 155–160.

Nowell, C., and T. Grijalva (2011): "Trends in co-authorship in economics since 1985," *Applied Economics*, 43(28), 4369–4375.

Seiler, C., and K. Wohlrabe (2012): "Ranking economists on the basis of many indicators: An alternative approach using RePEc data," *Journal of Informetrics*, 6(2), 389–402.

Sutter, M., and M. Kocher (2004): "Patterns of co-authorship among economics departments in the USA," *Applied Economics*, 36(4), 327–333.

Zimmermann, C. (2013): "Academic rankings with RePEc," *Econometrics*, 1(3), 249–280.

Table 2: Descriptive Statistics

Variable name	**Description**	**Min**	**Max**	**Mean**	**SD**
year	Year = year date - 1990	1	23	16.00	6.00
year2	Year^2	1	529	292.23	167.71
authors	Number of authors	1	49	1.89	1.05
citations	Number of citations	0	3489	7.69	34.03
IF	Journal impact factor in RePEc	0	88	4.96	8.08
title	Title length (number of characters)	1	468	72.70	29.27
pages	Number of pages	1	100	17.53	9.53
SUMJEL	Sum of JEL categories of paper	1	13	1.81	0.90
JEL_A	General Economics and Teaching	0	1	0.02	0.15
JEL_B	History of Economic Thought, Methodology, and Heterodox Approaches	0	1	0.02	0.16
JEL_C	Mathematical and Quantitative Methods	0	1	0.13	0.33
JEL_D	Microeconomics	0	1	0.19	0.39
JEL_E	Mactroeconomics and Monetary Economics	0	1	0.14	0.34
JEL_F	International Economics	0	1	0.14	0.34
JEL_G	Financial Economics	0	1	0.17	0.38
JEL_H	Public Economics	0	1	0.09	0.28
JEL_I	Health, Education, and Welfare	0	1	0.09	0.28
JEL_J	Labor and Demographic Economics	0	1	0.12	0.33
JEL_K	Law and Economics	0	1	0.03	0.17
JEL_L	Industrial Organization	0	1	0.16	0.37
JEL_M	Business Administration and Business Economics, Marketing, Accounting	0	1	0.08	0.27
JEL_N	Economic History	0	1	0.02	0.14
JEL_O	Economic Development, Technological Change, and Growth	0	1	0.18	0.38
JEL_P	Economic Systems	0	1	0.05	0.21
JEL_Q	Agricultural and Natural Resource Economics, Environmental and Ecological Economics	0	1	0.11	0.31
JEL_R	Urban, Rural, Regional, Real Estate, and Transportation Economics	0	1	0.09	0.28

Table 3: Results for Poisson Regression

	Model 1		Model 2		Model 3	
Variable	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
Constant	0.347	0.000	0.344	0.000	0.206	0.000
year	0.013	0.000	0.013	0.000	0.011	0.000
year2	0.000	0.478	0.000	0.232	0.000	0.058
IF	0.003	0.000	0.003	0.000	0.003	0.000
SUMJEL	0.025	0.000	0.025	0.000		
citations			0.000	0.000	0.000	0.000
title					0.001	0.000
pages					0.003	0.000
JEL_A					-0.170	0.000
JEL_B					-0.286	0.000
JEL_C					0.093	0.000
JEL_D					0.023	0.000
JEL_E					-0.038	0.000
JEL_F					-0.020	0.000
JEL_G					0.061	0.000
JEL_H					-0.029	0.000
JEL_I					0.164	0.000
JEL_J					-0.003	0.585
JEL_K					-0.136	0.000
JEL_L					0.035	0.000
JEL_M					0.082	0.000
JEL_N					-0.154	0.000
JEL_O					-0.016	0.000
JEL_P					-0.002	0.782
JEL_Q					0.184	0.000
JEL_R					0.044	0.000
χ^2 (Significance)	3381.14 (0.00)		3517.35 (0.00)		9508.47 (0.00)	

The dependent variable in each model is the number of authors.